APPENDIX C

THE SIX SIGMA RULE

C-1. Introduction. In reliability analyses the random variables are represented by a probability density function. There are many probability density functions; however, for geotechnical reliability analyses the ones most commonly used are the uniform distribution, the triangular distribution, the normal distribution, and the lognormal distribution. One method of reliability analysis used by the Corps of Engineers is the reliability index method. This method is a first order second moment method of analysis. This means that only the first two moments (mean and variance) are used to represent the probability density function in the reliability analysis and all higher order terms are neglected in the Taylor Series expansion used to estimate the mean and variance of the performance function (natural log of the factor of safety). So when using the reliability index method no knowledge of the exact probability density function that represents a random variable is needed. Only the mean and the variance of the random variable are needed. The mean is typically called the expected value and the variance is equal to the standard deviation squared (σ^2). The expected value of a random variable is easily estimated by geotechnical engineers as that is the average value of the parameter or what one would expect the value of the parameter to be. The standard deviation is not as easy to estimate; however, several ways to determine the standard deviation are given in Appendix D. This discussion will deal with one of those methods for determining the standard deviation, the six sigma rule and variations of that procedure.

C-2. Six Sigma Rule. The six sigma rule is based on the normal distribution probability density function, see Figure C-1. The six sigma rule makes use of the experience of geotechnical engineers in estimating the value of the standard deviation. The six sigma rule is given by Equation 1.

$$\sigma = \frac{Largest\ \ Possible\ \ Value - Smallest\ \ Possible\ \ Value}{6} \tag{1}$$

By examining the normal probability density function in Figure C-1, the reason why the six sigma rule works becomes clear. Figure C-1 shows a normal probability density function with the percentage of the area under the curve for the mean plus or minus one standard deviation (a range of 2σ), plus or minus two standard deviations (a range of 4σ), plus or minus two and one half standard deviations (a range of 5σ), and plus or minus three standard deviations (a range of 6σ). For the mean plus or minus three standard deviations, 99.73 percent of the area under the normal distribution is included. Therefore 99.73 percent of all possible values of the random variable are included in this range. So, for a range of six sigma (the highest possible value to the lowest possible value of the random variable), essentially all the values represented by the normal distribution curve are included, thus the name the six sigma rule. In the literature, Duncan (1999) and Dai and Wang (1992) and in this document, the six sigma rule is sometimes called the three sigma rule because the six sigma range of the data is represented by a plus or minus three sigma.

C-3. Other Bounds of the Data. Other variation of the six sigma rule exist depending on your confidence in estimating the upper and lower limits of the values represented by the random variable. EC 1105-2-205, Risk-Based Analysis for Evaluation of Hydrology/Hydraulics and Economics in Flood Damage Reduction Studies, uses a four sigma rule and a two sigma rule to estimate the standard deviation.

a. The four sigma rule is given by Equation 2.

$$\sigma = \frac{E_{mean}}{4} \tag{2}$$

where: E_{mean} = the difference between reasonable upper and lower limits which bound 95 percent of all data.

b. The two sigma rule is given by Equation 3.

$$\sigma = \frac{E_{majority}}{2} \tag{3}$$

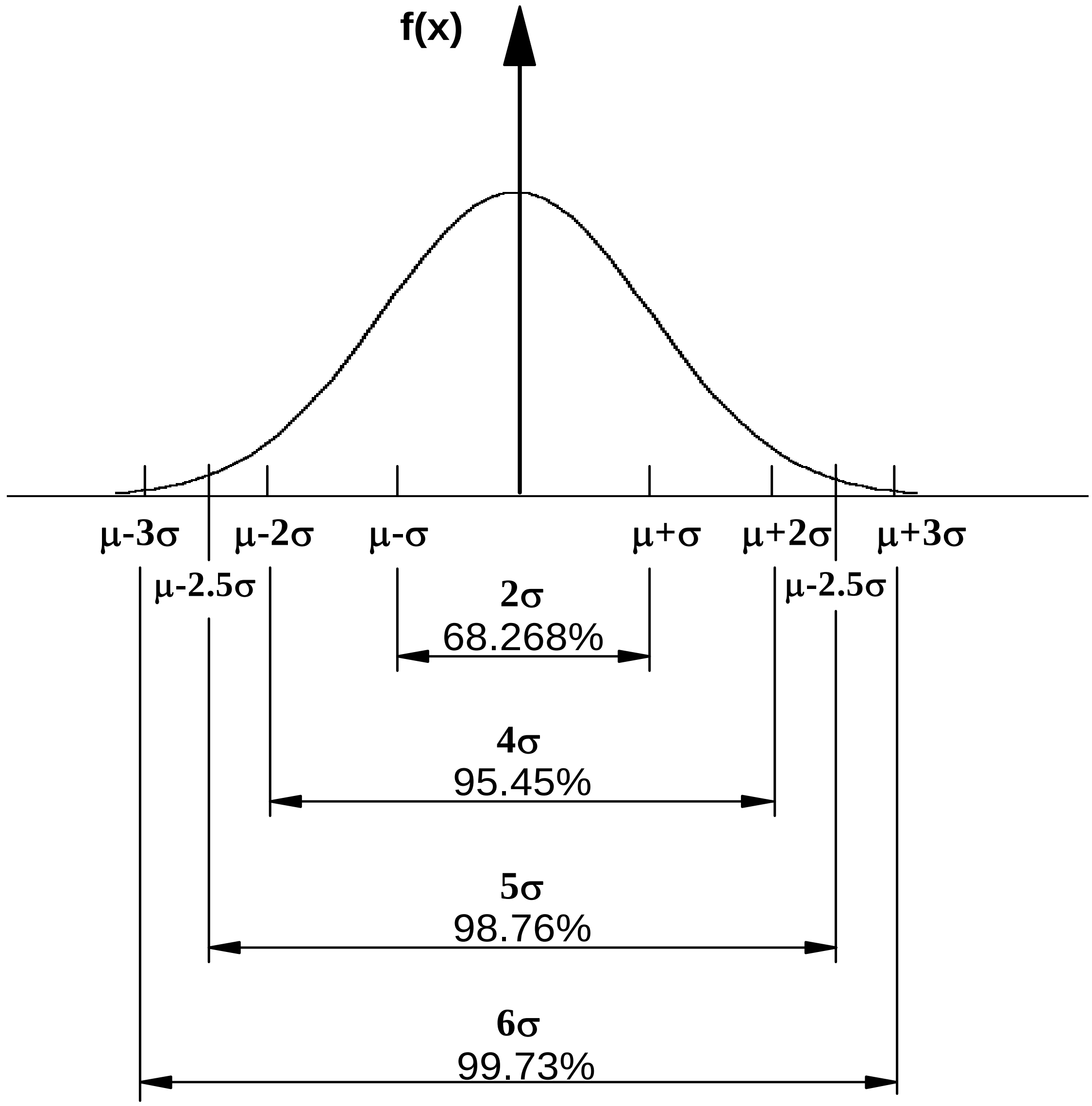
where: $E_{majority}$ = the difference between the upper and lower limits for the majority of the data bounding 68 percent of all data.

c. The five sigma rule is given by Equation 4.

$$\sigma = \frac{Largest\ Reasonable\, Value - Smallest\ Reasonable\, Value}{5} \tag{4}$$

where the difference between the largest reasonable value and the smallest reasonable value of the random variable includes 99 percent of all data.

C-4. Summary. All of the sigma rules are based on the normal distribution shown in Figure C-1 and the confidence the geotechnical engineer has in estimating the upper and lower limits of the data range representing the random variable. If there is 100 percent confidence in estimating the upper and lower limits of the random variable, use the six sigma rule. If there is a 99 percent confidence in the estimating the upper and lower limits of the random variable, use the five sigma rule. Likewise for 95 percent confidence in the estimating the range of the random variable, use the four sigma rule; and for 68 percent confident in the estimating the data range, use the two sigma rule. From a practical stand point, it would be difficult to justify using the two sigma rule since it would not be easy to estimate an upper and lower bound on 68 percent of the data. However, one could easily justify use of either the six, five, or four sigma rules and EC 1105-2-205 recommends hydrology/hydraulic engineers use the four sigma rule for calculating the standard deviation for water level data.



NORMAL DISTRIBUTION

Figure C-1